St Analysis Division
Division

Division de l'Analyse Structurelle

(2)

The Process Encyclopedia

Introduction and Feasibility

Statistics Canada
Statistique Canada

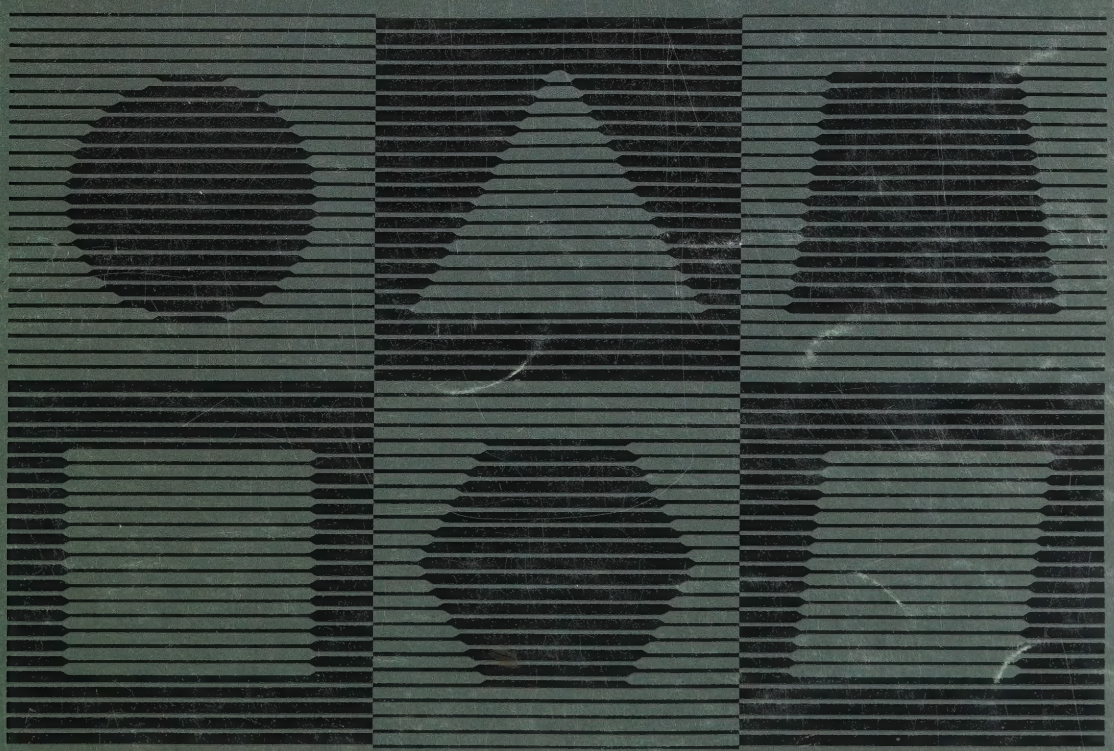
Structural Analysis Division

Division de l'Analyse Structurelle

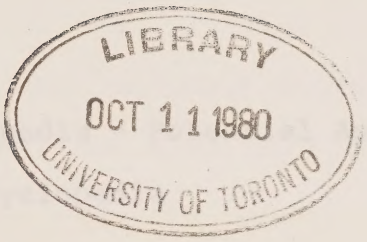
(2)

The Process Encyclopedia

Introduction and Feasibility



78-10-05



The Process Encyclopedia

Introduction and Feasibility

B.C. McInnis

William S. Page

R. Hoffman

Structural Analysis Div.

Table of Contents

1) Introduction

It has become apparent that physcial process information is a valuable source of data for extended economic analysis. In this paper we will make clear in what sense we speak of extended economic analysis and what is meant by physical process information. We will also outline the design of a means for collecting, editing, storing and retrieving this information in a manner appropriate to its intended use. A detailed design paper and papers dealing with specific applications of the process encyclopedia will be available at a later date.

We have used the term "extended economic analysis" to refer to the developing applications of traditional economic theory and extensions of the economists' methodologies to analysis and planning of natural resource systems (energy systems), ecological systems and the explicit representation of engineering technologies.

In particular we are referring to the general class of structural economic models originating with the input-output model of Leontief[1], generalized to rectangular industry by commodity input-output models by Stone[2], Gigantes[3] and Matuszewski[4], and incorporated in the wider theory of activity analysis by Koopmans[5]. The framework of activity analysis has been applied to the development of numerous engineering process models such as those employed in resource and energy analysis and

to the field of engineering economics in general[6]. This development has culminated in a synthesis based on the physical principles of the conservation of mass and energy, and is referred to as the Materials Energy Balance Statistical Subsystem[7] (MEBSS). The general aim of MEBSS is to facilitate analysis of alternate materials policy at the national level.

Physical process information refers to information available prior to economic analysis regarding the material transformation processes operative or potentially operative internal (and perhaps external) to the economy. The sources of this information are the theoretical investigations of physicists, chemists, and biologists and the accumulated practical knowledge of the related engineering disciplines. As will be discussed later, physical process information can be quantified on the basis of "unit processes".

The process encyclopedia, then, entails the identification of a taxonomy of processes and the quantification of processes in terms of the estimation of the parameters that describe unit processes. The information contained in the process encyclopedia will be semi-structured in the sense of being both quantitative and qualitative.

2) Why a Process Encyclopedia - Potential Applications

The process encyclopedia will serve primarily as a tool

underlying the construction of extended economic simulation models for support of policy analysis and planning. It will have obvious applications to the systems identification problem in providing the model designer (systems analyst) with a large repertoire of a priori structural information regarding possible system components. This type of information is invaluable whenever the method of analysis of a system is analogous to the so-called systems approach, in particular when it is possible, at least in part, to carry-out the analysis by the identification and simulation of the underlying material transformation and flow processes.

As examples of the above class of models we may cite literature in the areas of environmental modelling, natural resource and energy analysis, long term macro-economic modelling, micro-economic analysis and operations research. Appendix I contains a selected summary of such research and literature.

As mentioned above, it has been proposed that a system of detailed statistics be developed based on the conservation identities of mass and energy. This system of statistics (MEBSS) would be compatible with and would interface with the economic system of national accounts. MEBSS would be primarily concerned with the support of environmental and resource related studies and is intended to address questions of materials utilization policy. The process encyclopedia as proposed in this paper would serve as a major component of the MEBSS.

Depending on the level of detail present, the process encyclopedia has applications as a common repository of technical information in aid of research, development and design of industrial processes. On an international basis the process encyclopedia could serve as a vehicle for disseminating technological information and promoting the cross-fertilization of technological research and development programs.

3) What is a Process Encyclopedia

3.1 General Description

The process encyclopedia is a database containing semi-structured information to be used primarily in support of process-oriented economic and extended economic modelling, and the development of frameworks with particular emphasis on the representation of technology. As mentioned above the process encyclopedia also has important applications to less structured problems in the areas of decision making (policy analysis) in general and in the dissemination of technological information. The process encyclopedia presents an annotated taxonomy of industrial (and perhaps more general) processes.

3.2 Characteristics

In considering the above applications we discern that the following characteristics are highly desirable for the process

encyclopedia:

a) internal consistency of the data

The process encyclopedia may be employed as a benchmark or yardstick against which one may compare the parameter estimations of diverse models. It has not been infrequent in the past to observe deviations in the basic parameters of models, which purport to deal with the same subject, and thereby yield vastly different results. In reconciling such models one must face the problem of consistency again and again.

b) standardization of terminology, classifications and units of measure

In dealing with disaggregate, system-wide models and frameworks a great deal of effort should be placed on reducing unnecessary confusion. The communication between the diverse specialists who may be involved in the development and use of large scale models and frameworks presents special problems not the least of which is the development of a common vocabulary.

c) variable level of detail

In order to help bridge the gap between the detailed data requirements of process engineers and the high level aggregate information requirements of the economist and planner it is

necessary to retrieve information from the process encyclopedia at a user-selectable level of detail. The data requirements of a given model could have varying levels of detail in different sectors of the model. In achieving this variable level of detail it is not adequate to allow the retrieval to add apples to railway ties, but rather the process encyclopedia must allow one to abstract apples and oranges as fruit, and railway ties and two-by-fours as wood products. Moreover, a given telescoping hierarchical classification is not in general sufficient, therefore the underlying structure of the process encyclopedia must be more general (i.e. classification systems that are isomorphic to cascade graphs as opposed to tree graphs).

3.3 Data Collection

The structural information contained in the process encyclopedia will make it possible to collect much more detailed, selective, and accurate information regarding installed capacity and production statistics than is currently provided by surveys such as a manufacturing census. A particular firm or establishment could be characterized as a collection of processes. Detailed establishment-specific questionaires could be developed based on the known process composition of a given establishment. Unit process parameters contained in the encyclopedia could serve as editing and consistency checking base-lines to insure that the observed data does not deviate substantially from the known operating characteristics of the

processes employed by the establishment. In this way it will be possible to collect information with a level of accuracy that has not been technically possible before. Physical engineering parameters will make it possible to imput significant "unobservable" data (data which, in fact, may not even be collected or known by the managers of the establishment such as pollution emission levels etc).

Data collection is a two-way street. Not only will the process encyclopedia enable more accurate and complete data collection, but the data collection activity itself may reveal underlying processes. This, in turn, will make it possible to identify the detailed components of technological change i.e. technology substitution versus internal technological change and to quantify these changes. We should note, as an aside here, that this use of the process encyclopedia represents the imposition of detailed accounting principles, which could be argued philosophically to be the fundamental characteristic of systems identification. This concept of imprinting as a method for describing the production system is not new to economics. It has, however, recently become more prominent via the methodology of energy analysis and resource management.

A complete treatment of the dynamic behaviour of the economic production system also requires the representation of capital creation or investment. Thus the process encyclopedia must include detailed unit capital-bundle requirements of the

processes as well as the unit input-output flows. An interesting alternative is the inclusion of capacity creation processes. This allows a uniform treatment of unit-capacity as a production good distinguished by the fact that unit-capacity as a good is not consumed by any other process except the associated capacity ageing process. The unit capital-bundle for a particular process, then, is the unit input to the associated capacity creation process.

4) What Constitutes a Process?

The central concept is that of a process. The concept of a process is very nearly elementary (a priori or axiomatic) and therefore rather difficult to define rigourously; in fact, it is perhaps best understood in reference to the contexts in which it is used. We will, however, attempt to place the concept in perspective.

Ayres[8] defines a set of 11 elementary "micro-processes" involving physical and chemical changes. Each of the 11 are further broken down according to the detailed change taking place. This classification is illustrated in table 1 below. The chemical engineering terminology of unit-operations and unit-processes is the basis for this classification. Ayres claims that all processes "in the large" or macro-processes can be analysed as compositions of the defined set of elementary micro-processes and goes on to demonstrate how the a priori

compatibility of micro-processes and elementary materials (as classified by characteristics) induces a structural ontology greatly simplifying this analysis. The economic entity termed a firm, establishment, plant, or facility can be described, in turn, as being a composition of macro-processes, i.e. itself a macro-process.

Table 1

P1	Transportation (solids or fluids)
P2	Change of energy state (solid or fluid)
P3	Change in physical form (mainly solids)
P4	Physical integration (solids or mesh)
P5	Physical disintegration (solids or mesh)
P6	Physical association
P7	Physical dissociation or separation
P8	Surface treatment or finishing (solids)
P9	Chemical dissociation or decomposition
P10	Chemical association or synthesis
P11	Isomerization

As the process encyclopedia is aimed primarily at servicing economic model building and framework development, information in the encyclopedia will be concentrated primarily at the level of the "economic unit process" i.e. the lowest level (most elementary) macro-process which has potential economic (monetary)

significance. A facility has been analysed into its constituent economic unit processes when we can no longer define any smaller subset of processes which have a unique economic identity in the sense that the material inputs and outputs of the processes potentially have a definable market.

In the process encyclopedia economic unit processes will be described in both qualitative and quantitative terms. Quantitative information will identify and provide estimates of commodity input requirements and commodity outputs of each unit process. Engineering design capacity figures (theoretical limits) and estimates of actual load factors (operating characteristics) will quantify the expected production or activity level of each process.

Thus, in terms of the following expression

$$x_k \, a_{nk} \quad (n=1,2,\ldots N)$$

the process encyclopedia will contain information on the expected flow of commodity n per unit time input to ($a_{nk} < 0$) or output from ($a_{nk} > 0$) the k-th process; where a_{nk} are the unit process parameters and x_k is the activity level which includes the concept of a load factor. No summation is implied. Note that this formulation is for illustrative purposes only, the exact

form of the quantitative information not yet being determined.

5) Feasibility and Cost

Although legitimate estimation of feasibility and cost can only be made after a pilot study has been undertaken, there are a number of comments and observations that are in order and which indicate the reasons for our initial feeling that the process encyclopedia is feasible.

First, some estimate of expected ultimate size would place the encyclopedia in perspective. It is obvious that a considerable degree of disaggregation is required at the level of basic processes. It is our estimate that there is, for our purposes, on the order of 5000 significant economic unit processes operative in industrialized economies of today. Further, since the commodity flows are specified in quantity (as opposed to value) terms a great degree of disaggregation is implied in the classification of commodities as well, perhaps as many as 2000 categories. Semi-quantitative and qualitative notes and comments will be associated with each process where the available information does not easily fit the input-output unit-capacity framework. It is important to note the strong emphasis on quantity as opposed to financial data. In practice, it will be necessary in some cases, due to lack of data and difficulty of quantification, to use monetary figures as proxies for the underlying commodity flows. The process encyclopedia,

then, will contain information on "paradigm processes" as opposed to real observations. Observations of processes in place and many other sources of information will serve, in the construction and updating of the encyclopedia, as sources of calibration of these paradigm processes.

Structural information concerning the composition of aggregate processes in terms of the sub-processes involved will also be contained in the encyclopedia. Maintaining consistent data regarding each process and sub-process will be the major means by which the process encyclopedia will help to bridge the gap between micro-engineerng analysis and macro-economic analysis. Support of this type of information implies a thesaurus-like network structure. The interface to the database must allow the retriever of information to navigate up or down levels of abstraction and aggregation as well as across a particular level.

The computer capability, in the form of data base software, that can handle this type and size of structure is extant and no new developments are required. However, considerable adaptation of existing systems may be needed to match our needs.

The existence and collectability of the data can be imputed by the existence of numerous process encyclopedias and process data bases[9] for specific engineering industrial applications. Moreover our encyclopedia may draw heavily on these extant

sources.

Our experience in maintaining large data bases indicates that the resource requirements of maintaining and providing such an encyclopedia is not prohibitive. We would anticipate requirements on the order of 3 person-years.

The pilot development of the encyclopedia requires a threshold amount of information to be collected and a major portion of the software to be developed. We anticipate manpower requirements on the order of 10 person-years to reach a level where a reasonably credible estimate of ultimate feasibility can be made.

6) The Process Encyclopedia Information system

The process encyclopedia will also place a strong emphasis on documenting the sources of the data contained within it. The sources of data are many, of differing levels of reliability and of differing types (e.g. monetary and physical data) as Appendix II points out, including physical laws of conservation of mass and energy, administrative data, both government and private survey data, engineering studies, and casual estimates by experts.

The integration of these sources of data into a standard form will require a number of information tools to aid the contributor

of data. These will include the access to and ability to manipulate different classification schemes of processes, industries, commodities and materials, the ability to aggregate, disaggregate and extend these classifications, and to construct new classification schemes. A time series of commodity and price information will often be necessary to relate physical and monetary units. Conversion information between differing physical units of measure will also be necessary. In what follows we will refer to the total system of contribution and retrieval aids, in addition to the process encyclopedia proper, as the process encyclopedia information system.

Since one of the goals of the process encyclopedia is to promote comparability and standardization of data used in diverse studies it is of fundamental importance to establish the credibility of its contents. This may be partly achieved by detailed documentation of sources. Documentation of data sources will be important in at least two other respects. First, the development of the process encyclopedia will require continuing input and data massaging by many people with very different backgrounds over a fairly long period of time. Documentation will help to integrate these contacts with the process encyclopedia into a manageable framework which will serve to help co-ordinate the continued development of the encyclopedia. Secondly, in its capacity as a tool for disseminating technological information, it may be frequently valuable for researchers to have recourse to the original studies for the

necessary details. Thus, the process encyclopedia has the function of a technical literature information system as well. As such, considerable advantage may be gained in defining how the process encyclopedia will interface with other existing technical literature information systems. However, we will delay further discussion of this until a later paper.

It is the hope, of course, of the designers of the process encyclopedia that it be used by and of use to a large number of organizations for a large number of studies and that it will continue to be useful in the future.

Toward this end two provisions are important:

a) Use Trace Facility - within the process encyclopedia information system provision must be made for tracking the use of the information contained in it. As new studies are completed and documents generated with references to it, an up-to-date list of citations that are context keyworded should be maintined. This should provide a guide to future development.

b) User Specified Retrieval Quality Control Facility - within the system data entered should be classed by source and provision made in the retrieval system for screening the data based on user determined descriptors of source.

7) Summary

A process encyclopedia is a structured data base containing information on "economic processes" that allows decentralized data collection, user controlled data retrieval and "paradigm process generation".

In its ultimate form it would contain some 5000 processes dealing with some 2000 commodity groups.

It would support various model and framework developments ing.

Based on our perception of existing analytic requirements of both governments and industry, our experience with existing computer technology and our knowledge of existing data sources, we think the concept of a process encyclopedia is ripe for being made operational and applied.

References

1) Leontieff, W., 1951 - "The Structure of the American Economy 1919 - 1939", New York, Oxford University Press Ed 2.

2) Stone, R. (Ed) - "A Programme for Growth," Volumes 8, 9, 10 Chapman and Hall, London (Vol. 8), 1968; (Vols. 9 and 10), 1970. Also See: United Nations: Department of Economic and Social Affairs; Statistical Office of the United Nations, A System of National Accounts, Studief the American Economy 1919 - 1939", New York, Oxford University Press Ed 2.

2) Stone, R. (Ed) - "A Programme for Growth," Volumes 8, 9, 10 Chapman and Hall, London (Vol. 8), 1968; (Vols. 9 and 10), 1970. Also See: United Nations: Department of Economic and Social Affairs; Statistical Office of the United Nations, A System of National Accounts, Studies in Methods, Series F, No. 2, Rev. 3, New York: 1968.

3) Gigantes, T., "The Representation of Technology In Input-Output Systems, In: A.P. Carter and A. Brody (eds), "Contributions to Input-Output Analysis" North Holland, Amsterdam, 1970.

4) Matuszewski, T., "Partly Disaggregated Rectangular Input-Output Models and Their Use for the Purposes of a Large Corporation", In: A.P. Carter and A. Brody (eds) "Input-Output Techniques", North Holland, Amsterdam, 1972.

5) Koopmans, Tjalling C., (ed), Activity Analysis of Production and Allocation, Cowles Foundation Monograph 13, Yale University Press, New Haven, 1951.

6) The literature that can be cited in this area is voluminous and a discussion of selected parts of it can be found in Appendix I of this paper. The kind of work we refer to is perhaps best illustrated in;
Ayres, R., et. al., Materials Process Product Model, an International Research and Technology Co. report IRT-305-FR prepared for National Science Foundation under contract NSF-C652.

7) U.N. Statistical Office, Draft Guidlines for Statistics on Materials/Energy Balances, draft working paper prepared for the U.N. Statistical Office Ref. No. E/CN.3/492, 1976.

8) Ayres, R., A Materials Process Product Model in "Environmental Quality Analysis", Kneese and Bower (eds) John Hopkins Press, Baltimore, 1972.

9) We can cite for example; Kirk, R.E., and Othmer, D.F., "Encyclopedia of Chemical Technology", Vol. 1-25, John Wiley & Sons, New York, 1963-72; which is public and the "Process

Encyclopedia" maintained by the Stanford Research Institute which is proprietary. Special purpose encyclopedias or data bases exist for commercial use. We can cite for example the materials properties data base maintained by SACDA (Faculty of Engineering Science, University of Western Ontario, London, Ontario) to design special distillation processes.

Appendix I

Selected Literature Review

Process information used to support analysis appears in literature of all types, from published books and papers, through consulting reports requested by governments, industries and non-profit organizations to in-house management reports in all forms of institutions. Accessability to this information decreases from good in the first mentioned category to poor in the last. Because of this the need for process information of the type discussed in the main body of the paper is likely to be underestimated. The intention of this appendix is to give the interested reader literature access doors and to create a sense of the type of use of process information that is extant. We arbitrarily divide the universe into three parts, General Socio-Economic Analysis, Environmental Analysis, and Material Resource Analysis.

I.1) General Socio-Economic Analysis

I.1.1) Macro-Analysis

Any macro-analysis that deals with production system compositional effects and the relation of economic activity and resources requires some form of production system or function. Process information is the fundamental informational support of such production systems or functions and as such is basic to all such analysis. As an example consider:

I.1.1.1) Almon, C., Jr., et. al. 1985: Interindustry Forecasts of the American Economy, Lexington Books, Lexington Mass., 1975.

In this medium term scenario analysis of the U.S. economy the basic production system is a transaction input-output system. This is effectively a high level aggregation of processes. Moreover, specific process information is considered in projecting coefficient evolution (See page 157). The applications section of the book, chapter 10, is of particular interest because of its use of the concept of a "skirt" (Page 188). The development of "skirts" provides a natural process encyclopedia application.

I.1.2) Micro-Analysis

In this area much of the analysis only implies a reliance on process information. This is most easily seen in areas where capital costs are specified as a single monetary figure. If however the costs were specified as a commodity bundle, which is implicit on the encyclopedia, sensitivity analysis with

respect to capital costs would take on a whole new dimension. Consider for example:

I.1.2.1) Rowse, J., Toward Optimal Capacity Expansion for an Electric Utility: the case of Saskatchewan, Canadian Journal of Economics, XI, Number 3, 1978.

If the cost functions were of the commodity bundle type which would be quantified by the process encyclopedia, then sensitivity to relative commodity price variations could be carried out in a controlled and integrated manner.

The use of process information dominates investment decision analysis as evidenced by reference to the engineering literature for example:

I.1.2.2) Peters, M.S. and Timmerhaus, K.D., Plant Design and Economics for Chemical Engineers, McGraw Hall, New York, 1968.
and
I.1.2.3) Manne, A.S., Investment for Capacity Expansion: Size, Location and Time-Phasing, George Allen and Unwin Ltd., London, 1967.

It is also of importance for the analysis of technological substitution as exemplified by

I.1.2.4) Ayres, R. et. al., Materials-Process-Product Model (See reference section of this paper for citation).

I.2) Environmental Resource Analysis

The use of process information to relate economic activity to the use of environmental cleansing processes is exemplified by:

I.2.1) Russel, C.S., Residuals Management in Industry: A Case Study of Petroleum Refining, John Hopkins University Press, London, 1973.

I.2.2) Russel, C.S. and Spofford, W.O., Jr., A Quantitative Framework for Residuals Management Decisions, in "Environmental Quality Analysis" Eds. Kneese A.V. and Bower B.T., Johns Hopkins University Press, London, 1972.

I.2.3) Victor P.A., Pollution: Economy and Environment, University of Toronto Press, 1972.

I.3) Material Resource Analysis

The use of process information to relate economic activity to material resource use can be found exemplified in:

I.3.1) Ayres R. et. al. Materials-Process Product Model (See reference section of this paper for citation).

I.3.2) Bower B.T., The Economics of Industrial Water Utilization, in "Water Research",eds. Kneese A.V. and Smith S.C., Hopkins University Press, London, 1966.

I.3.3) Winstanley G. and Emmet B., Energy Requirements Associated with Selected Canadian Energy Developments, Research Report No. 13, March, 1977, Energy Mines and Resources, Office of Energy Conservation.

I.3.4) IBI Group, Indirect Energy in Transportation, March 1978, Strategic Studies Branch, Transport Canada.

Numerous other citations are possible including an annotated list of large model developments that could use a process encyclopedia. It is planned that this annotated citation list will be developed over the next year as part of the process encyclopedia project.

Appendix II

Types and Sources of Data

II.1) Types of Data

It is convenient to class data into four types. These are listed below in order of decreasing reliability.

II.1.1) Industrial Scale Process Data

In this category we have observations on processes in plants that are operating in the economy now.

II.1.2) Pilot Scale Process Data

In this category are observations on processes in plants that are operating but are of such a scale that upscaling may cause significant variations in the process data. However, the observations are on processes that are physically feasible as are the ones in the first category.

II.1.3) Prototype Scale Process Data

In this category we have observations on processes that are operating only in an embryonic state. Thus the core of the process is physically feasible but the ancillary "life support" appendages are presently artificially provided. Scale up to industrial scale may result in operating variation that induces the process to be non-economic.

II.1.4) Hypothetical Process Data

In this category are placed expert opinion, based on scientific principles, on the nature of processes yet to be shown to be physically feasible.

II.2 Sources of Data

It is useful to distinguish five main sources.

II.2.1) Private Sector Company Records

This source of information has constraints of a proprietary nature, however it is the most reliable and robust source. Two sub-types of records should be distinguished, those on operating plants and those on capital composition. To a large extent those two sub-types are kept by different organizations. The capital composition is largely known only to engineering consulting firms, while all firms have data on process operation.

II.2.2) Public Sector Agency Records

In many instances technologies have been developed by the public sector and records of the operating and capital cost structures are held by those responsible agencies.

II.2.3) Government Data Bases

There are two types of data bases that are of use. These may be called administrative and analytic.

II.2.3.1) Analytic

In this case we have the censuses carried out by various government bodies for the purposes of providing analytic tools for policy analysis.

II.2.3.2) Administrative

In order to administer certain laws, special purpose data bases are sometimes required and many of these contain useful partial information on processes.

II.2.4) Engineering Literature

A large amount of data of all data types can be found in the engineering literature of journals, trade publications, in-house working papers etc.

II.2.5) Expert Opinion

For various institutional reasons there is also a large amount of process information that is not published but is available through consultation with engineering experts.